

Amendments to the Specification:

Please amend the specification to include the enclosed Sequence Listing.

Please replace the paragraph beginning on page 18, line 8, with the following amended paragraph:

Fig. 10 lists the genetic identifiers of 49 proteinase K homologs obtained by BLAST searching of ~~Genbank~~ GENBANK.

Please replace the paragraph beginning on page 34, line 12, with the following amended paragraph:

In the example of a proline endopeptidase (~~Genbank~~ GENBANK A38086) there are many homologs and structures of homologs available. A detailed evaluation of various substitutions using three different methods 130 identified substitution F416Y as favorable. The scores from the various methods 30 are (i) the scores derived from favorability based on natural occurring substitutions using the PAM100 matrix is 5.29 (rank 1), (ii) the scores based on substitution found in a homolog expressed in the evolutionary distance of the homologs from the reference is 0.25 (rank 2), (iii) scores from positional variability of the sequence expressed in number of different types of amino acids found in that location is 3 (rank 7).

Please replace the paragraph beginning on page 52, line 31, with the following amended paragraph:

In the example of a proline endopeptidase (~~Genbank~~ GENBANK A38086) there are many homologs and structures of homologs available. Every possible substitution enumerated was assigned a score based on the PAM100 matrix. For example, substitutions for position 416: F416Y ranks number 1 and has a score of 5.24, F416L ranks 565 with a score of 1.2 and F416I ranks 1765 with a score of -0.83.

Please replace the paragraph beginning on page 74, line 18, with the following amended paragraph:

It will be appreciated by one skilled in the art that each different method for deriving relationships between biopolymer sequences and activities can differ in the precise values of their outputs. In some embodiments of the invention it is therefore desirable to combine the outputs from two or more such methods for subsequent uses. This corresponds to step 06 in Figure 2. There are a variety of ways in which such outputs can be combined. In some embodiments, each output can be independently applied to the subsequent design of biopolymer variants (Figure 2, step 07) or the modification of parameters or weights used by expert system 100 for the selection of substitutions (Figure 2 step 02) or the design of biopolymer variant sets (Figure 2 step 03). In some embodiments, average values (or some other mathematical function of two or more values derived by two or more sequence-activity models) can be calculated for the regression coefficient, weight or other value describing the relative or absolute contribution of each substitution or combination of substitutions to one or more activity of the biopolymer (*e.g.*, as defined in Equation 4 below). In some embodiments, the standard deviation, variance or other measure of the confidence with which the value describing the contribution of the substitution or combination of substitutions to one or more activity of the biopolymer can be assigned (*e.g.*, as defined in Equation 4 below). In some embodiments, the rank order of preferred substitutions is used to combine the methods. In some embodiments, the additive (linear variables) and non-additive components (non-linear variables) of each substitution or combination of substitutions is combined:

$$[(Eq. 6)] (Eq. 4) V_{ix} = f(M_1(i_x), M_2(i_x), \dots, M_j(i_x))$$

where,

V_{ix} is a combined measure of one of the descriptors measuring the performance of a biopolymer in which monomer x is substituted at position i ;

$M_j(i_x)$ is a measure of one of descriptors measuring the performance of a biopolymer in which monomer x is substituted at position i , determined by sequence-activity correlating method j ($M_j(i_x)$ is the contribution of i_x as determined by Model j); and

$f()$ is some mathematical function.

Please replace the paragraph beginning on page 111, line 30, with the following amended paragraph:

The proteinase K gene was used as probe against ~~GenBank~~ GENBANK using BLAST based algorithms. A BLAST score was chosen as a cut-off that identified more than ten but less than one hundred related sequences. This search identified the 49 sequences identified in Figure 10.

Please replace the paragraph beginning on page 113, line 14, with the following amended paragraph

The BLAST search of ~~Genbank~~ GENBANK for proteinase K homologs also revealed that proteinase K is homologous to subtilisin and other serine proteases. Subtilisin in particular has been extensively studied. The structures of naturally occurring and variant subtilsins have been obtained, and there is a large body of data regarding the functional effects of a substantial number of mutations. See, for example, Bryan, 2000, Biochim Biophys Acta 1543:203-222. Sequence and structural alignments of proteinase K with subtilisin allowed for the identification of homologous positions in proteinase K having changes known to improve activity or thermostabilize subtilisin. This information was incorporated into the knowledge base 108. This is an example of pre-processing information.

Please replace the paragraph beginning on page 122, line 25, with the following amended paragraph

The *Myxococcus xanthus* prolyl endopeptidase sequence used was the one defined by genetic identifier [gi:4838465] in ~~Genbank~~ GENBANK, and accessed by searching for this identifier using the NCBI browser. The following homologs were identified: gi|17131625; gi|24348832; gi|28808634; gi|6048357; gi|4973227; gi|28809898; gi|6460324; gi|216201; gi|27358772; gi|216707; gi|456523; gi|3805974; gi|21727153; gi|4529992; gi|148698; gi|19347837; gi|22946157; gi|11691900; gi|15277538; gi|6456472; gi|6561876; gi|5689035; gi|26343763; gi|5103285; gi|26345256; gi|21040382; gi|164621; gi|9971902; gi|558596; gi|3043760; gi|904214; gi|28502989; gi|17385666; gi|9558588; and gi|15291259.

Please replace the paragraph beginning on page 126, line 9, with the following amended paragraph

In this example, the optimization procedures of the present invention are illustrated for an antibody that binds and neutralizes Respiratory Syncytial Virus (RSV). The sequence of one such antibody is publicly available (~~Genbank~~ GENBANK accession # AAF21612). A significant benefit of the computational antibody design system using the methods described in this invention is that only relatively small numbers of variants need to be synthesized and tested. This allows the use of functional tests that are more comprehensive than binding assays. Viral neutralization for example, is an important antibody function but the sequence and structural determinants are poorly understood.